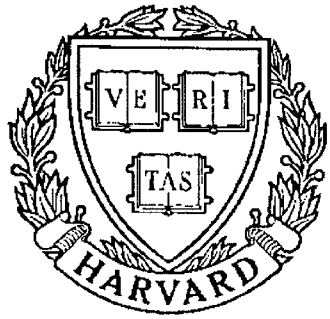
TECHNICAL RESEARCH REPORT

SYSTEMS RESEARCH CENTER



Supported by the National Science Foundation Engineering Research Center Program (NSFD CD 8803012), Industry and the University

Neural Network Solutions to Problems in "Early Taction"

by Y.C. Pati, P.S. Krishnaprasad and M.C. Peckerar

TR 89-104

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
1989		**00-00-1989 to 00-00-1989**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Neural Network Solutions to Problems in 'Early Taction'	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of Maryland,Systems Research Center,College Park,MD,20742**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
see report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**35**	

NEURAL NETWORK SOLUTIONS TO PROBLEMS IN 'EARLY TACTION'

Y. C. Pati*† P. S. Krishnaprasad* M. C. Peckerar†

*Systems Research Center, University of Maryland,
College Park, MD., 20742
†U.S. Naval Research Laboratories, Code 6804
Washington D.C., 20375

Abstract

In this paper we examine the application of artificial neural networks to low level processing of tactile sensory data. In analogy to the term *early vision*, we call the first level of processing required in tactile sensing *early taction*. Associated with almost all existing realizations of tactile sensors, are fundamental inverse problems which must be solved. Solutions to these inverse problems are computationally demanding. Among such inverse problems, is the problem of 'deblurring' or deconvolution of data provided by an array of tactile sensors which is also assumed to be corrupted by noise. We note that this inverse problem is ill-posed and that the technique of regularization may be used to obtain solutions. The theory of nonlinear electrical networks is utilized to describe energy functions for a class of nonlinear networks and to show that the equilibrium states of the proposed network correspond to regularized solutions of the deblurring problem. An entropy regularizer is incorporated into the energy function of the network for the recovery of normal stress distributions. It is demonstrated by means of both computer simulations and hardware prototypes that neural networks provide an elegant solution to the need for fast, local computation in tactile sensing. An integrated circuit prototype of the proposed network which has been designed and fabricated is discussed as well.

1 Introduction

Recent years have seen a significant increase in the complexity of tasks performed by robotic manipulators. As the complexity of these tasks continues to grow, the need for automated tactile sensing becomes increasingly evident. The term tactile sensing, as used here, refers to the continuous sensing of forces over regions of contact. Included in this definition of tactile sensing are the more rudimentary operations of binary contact sensing and pressure sensing.

Since the specific requirements of robot tactile sensing have not as yet been clearly defined, it is often useful to view tactile sensing in humans as a model for artificial tactile sensing. Tactile sensing in humans is a dynamic process in which dexterous hands are used in conjunction with dense arrays of subcutaneous sensors to extract information about the contact which is necessary for feature identification and formulation of manipulation strategies (see [13]). An 'ideal' artificial tactile sensing system should strive to perform a similar function.

The problem of tactile sensing can be hierarchically separated into three stages.

(a) At the lowest level of the hierarchy, there are the device level problems of designing a tactile sensory device, and of designing a dextrous manipulator to be equipped with such

sensors. According to a well known survey [18], tactile sensors should be distributed in arrays on thin, flexible, compliant substrates. Also, since tactile sensing is based upon physical contact, it is required that the entire structure comprising the tactile sensors be mechanically durable and robust against environmental variations. It was also suggested in [18] that a fair amount of preprocessing of the sensory data be performed in proximity to the sensor so as to limit the quantity of data to be transferred to the central processing unit.

(b) Given data from the tactile sensors, the second stage of the hierarchy is concerned with low-level processing and extraction of information. For example, often raw data provided by the sensors are not measurements of contact forces, but are related in some manner to the stress profile over the region of contact. In such instances, the inverse problem of determining the stress at the contact surfaces must be solved in order to obtain a meaningful interpretation of the sensory data. Detection of edges, extraction of geometric information and identification of conditions such as slippage are further examples of tasks to be performed at this level of the hierarchy. It is this level of tactile sensory data processing that we term *early taction.*

(c) The top-most level in this hierarchical picture of tactile sensing is the level at which decisions are made, based upon information provided by the two previous stages. Formulation of manipulation strategies is performed at this level. For example, if a condition of slippage has been recognized, a manipulation strategy should be formulated so as to alleviate this condition. Also at this level of processing, decisions are made regarding the identification of a grasped object based upon information about texture, shape (which could have been determined from knowledge of edges), and material (which may possibly be determined from thermal conductivity).

It is among the goals of this paper to take a step towards integrating the two lowest levels of this hierarchy.

Several aspects of tactile sensing are of a very similar nature as problems encountered in computational vision. Essential differences lie in the fact that unlike vision sensors, which are remotely located with respect to their target, tactile sensors are required to be in physical contact with their targets. The 'deblurring' problem considered in this paper is an example of a problem which also arises in computational vision. Edge detection, which is used in vision to determine boundaries within the visual field, is required in tactile sensing to identify physical edges of objects, and locate holes to determine shape. A similar analogy can be drawn between motion detection in vision and identifying slippage in tactile sensing.

The set of processes that recover physical attributes of visible three dimensional objects from two-dimensional visual (intensity) images, is collectively termed as *early vision.* In a similar manner, the elements of the second hierarchical level of tactile sensing may be collectively termed *early taction.* Early taction then can be defined as the set of processes that recover the three-dimensional attributes of an object and properties of the established contact, from two-dimensional arrays of sensor measurements. As yet, the set of problems of which early taction is comprised have not been clearly defined, but recovery of stress over the contact region, edge detection, and identification of slippage are likely among them. In the case of human tactile sensing, the total number of sensor cells in a fingertip area of 20×30mm can exceed 60,000. In [18] it is suggested that a typical array of 10×10 sensor elements per square inch should suffice for many applications of tactile sensing. Considering only this reduced number of 100 sensors, the quantity of data that must be transferred to the central processing unit, if

all processing is to be done there, is still prohibitive. It would not in general be acceptable to have a cable running from the hand to the CPU consisting of several hundred conductors. In addition to simply conveying the information to the central processing unit, there is also the actual processing task which would demand much of the available processing time. It is clear that a great deal of preprocessing should be performed at or near the sensor array itself. In doing so, a simpler representation or compression of the data may be obtained which could then be conveyed to the central processing unit. Demands upon the processing power of the CPU would then perhaps be limited to, at most, only the top level in the tactile sensing hierarchy. In humans, useful information is often acquired through active manipulation. A grasped object is often scraped, rolled, mutilated to determine its shape and other properties [18]. Therefore it is reasonable to assume that in robots, as in humans, static contact analysis together with dynamic analysis may prove to be useful. *This dynamical aspect of tactile sensing, together with the fact that tactile sensing is generally to be applied in a 'real time' environment, implies the need for fast processing of tactile sensory data.*

In this paper we discuss a framework within which at least some of the second level of tactile information processing may be performed. The approach taken is designed to meet the requirement of local fingertip processing as well as the demand for fast computation. The problem considered here is the determination of surface stress from an array of sensors which provides measurements of strain induced in a elastic medium by contact at the boundary. Since it can be assumed that the data provided by the sensors is corrupted by noise it is necessary to consider this additional aspect of the problem as well. In order to meet the requirements of fast local fingertip processing, the paradigm of neural networks is considered.

Inspired by the work of neurophysiologists, psychologists and other researchers on the mechanisms of computation, learning and memory in biological systems, artificial neural networks are an attempt to reproduce the computational efficiency observed in the nervous system. An artificial neural network may be defined as a highly interconnected network of simple processing units. The processing units themselves are rarely more than simple amplifiers (usually nonlinear amplifiers such as those with sigmoidal characteristics are used), yet neural networks have in many instances demonstrated an ability to solve complex problems. The computational power of artificial neural networks is embedded in the nature of connectivities between the processing units (or *neurons*) of which they are composed. Neural networks[1] are usually regarded as being comprised of layers of neurons and the interconnections among them. Associated with a connection (also called a *synapse*) between two neurons, say neuron i and neuron j, is a number w_{ij} called the *weight* of the connection (sometimes referred to as the *synaptic weight*) between neuron i and neuron j, which determines the effect that the output of neuron i has upon the activity of neuron j. For example, if the output of neuron i is v_i then the input to neuron j due to the connection of neuron i to neuron j is given by $w_{ij}v_i$. It is the connectivity profile (distribution of connection weights) which determines the computational task performed by any given network. The nature of computation in a neural network is both parallel and asynchronous.

The paradigm of neural networks also provides a formalism for the analog hardware implementation of inherently parallel algorithms. Biological neurons are often modeled as integrators (with the sum of all inputs to the neuron as the integrand) composed with output functions. In terms of analog circuits this corresponds to a simple RC integrator circuit followed by an amplifier with the desired characteristics. Connections between neurons can be implemented as resistors. If the output of neuron i is a voltage which is connected to the input of neuron j through a resistor R_{ij} then Ohm's Law dictates that the current input to neuron j due to the

[1]The term *neural networks* will be used in reference to artificial neural networks unless otherwise indicated

output v_i of neuron i is $R_{ij}v_i$. Hence the synaptic weight w_{ij} is simply $1/R_{ij}$. Hence having arrived at a neural network model for computation, implementation of the network as an electrical circuit is a natural extension. Advances in microelectronics have provided the technology required for implementations of 'small' neural networks. VLSI implementations of networks consisting of hundreds, thousands, even tens of thousands of neurons, should be possible with further advances in microelectronic technologies, but as yet are unrealizable due mostly to limitations in integrated circuit density and physical size. Integrated circuit implementation of neural networks to perform the low level processing tasks of the second level of hierarchy in tactile sensing would provide fast computation which can be performed in close physical proximity to the sensors. The resultant of processing performed by such a network can then be transmitted to the central processing unit for higher level interpretation (and decision making based upon the interpretation) using relatively few conductors. Such an approach would also result in lower computational demand upon the CPU.

In Section 2, a particular inverse problem which arises in the context of tactile sensing is introduced. The inverse problem is formulated as a variational principle and 'regularization' is used to compensate for the ill-posedness of the problem and provide for reliable computation in the presence of sensor noise.

In Section 3 some aspects of nonlinear RC electrical networks are discussed. It is shown that under certain conditions it is possible to guarantee stability and explicitly determine a strict Lyapunov function for such a network. Since such a Lyapunov function is minimized in the course of natural time evolution of the network, steady state outputs of the network correspond to the solutions of a minimization problem (defined by the Lyapunov function).

An analog neural network to solve the inverse problem of tactile sensing is described in Section 4. Using the results of Section 3, an energy function for the proposed network is determined and shown to be equivalent to the variational principle formulated in Section 2 for regularized solution of the inverse problem. Computer simulations of the proposed network are presented in order to evaluate performance of the network in the presence of noise. Experimental results from a prototype breadboard model of the network as well as a preliminary attempt at integrated circuit implementation of the network are also discussed.

2 Inverse Problems In Tactile Sensing

In this section we examine the fundamental inverse problem which arises in the context of robot tactile perception. Inverse problems of a similar nature also arise in the field of computational vision in early vision problems which have often been designated as inverse optics (see Poggio and Koch [41]). Hadamard [17] in 1923 defined a problem to be well-posed if: (1)The solution exists, (2)The solution is unique, and (3)The solution depends continuously on the initial data. If any of the above three criteria are not satisfied, the problem is said to be ill-posed. Most inverse problems which arise in physical settings (such as those in early vision) are ill-posed in the sense defined by Hadamard and thus require special techniques to solve.

Numerous tactile sensors have been designed and fabricated for use with robotic end-effectors. All such sensors have been based on the deformation of materials by contact forces and the measurement and interpretation of the deformation to determine the forces inducing it. Examples of some approaches to tactile sensor design are to be found in [8], [15], [7], [43], and [38]. In typical approaches to tactile sensor design, the transduction process is effected by materials such as piezo-electric polymers [8] and crystals, conductive rubber [8], and piezoresistive materials such as silicon [56]. In [38] a tactile sensor is described which is

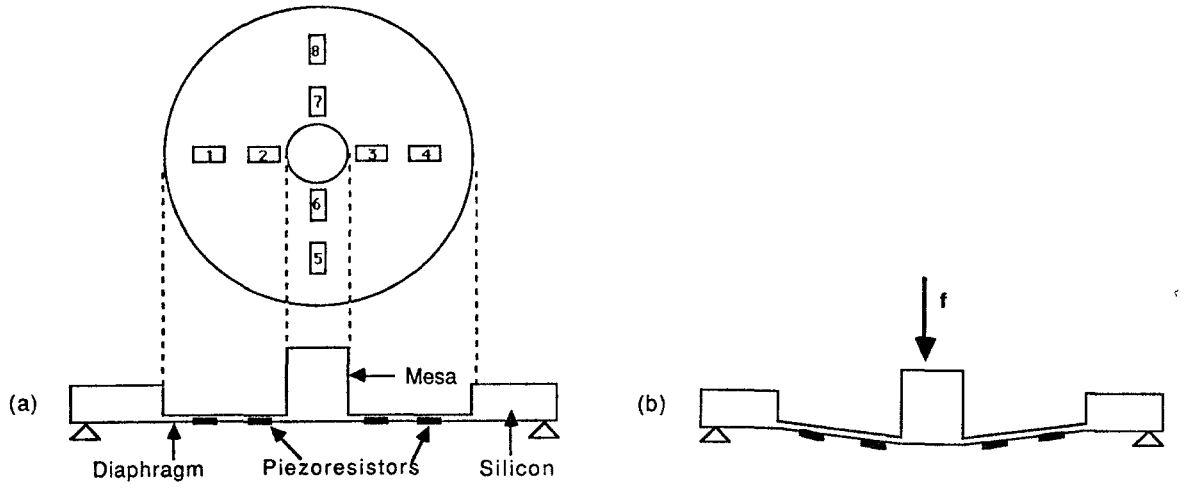


Figure 1: *(a) Schematic top and crossectional view of a single piezoresistive tactile sensing element. (b) Application of forces to the mesa deforms the diaphragm, causing changes in the resistance of the piezo resistors.*

designed to operate based upon changes in optical characteristics of a material at boundaries (total internal reflection). Such an approach has the advantage of very high spatial resolution since it is not necessary to construct arrays of discrete sensors. Another novel approach to tactile sensor design is described in [5] where transduction is based upon variations in magnetic fields which are measured by a VLSI array of Hall-effect sensors. In most approaches to tactile sensing, there arises an inverse problem, namely, given data from the sensors, determine the force profile at the contact region.

2.1 Tactile Sensors and Compliant Contact

For the purpose of providing a concrete example, we concern ourselves here with a silicon based piezoresistive triaxial tactile sensor with a compliant layer which has been designed and fabricated at the Naval Research Laboratories in Washington D.C. and described in [56].

The sensor is constructed as a micromachined silicon mesa surrounded by a thin diaphragm in which a number of diffused piezoresistors are placed (see Figure 1). An array of these sensors is constructed on a single die and then bonded and packaged to be mounted on a robot fingertip. Sensitivity of the tactile sensor is determined by the magnitude of strain induced in the resistors by a given force. In order to increase sensitivity of the sensors, it is necessary to decrease the thickness of the diaphragm and thereby increase the likelihood of its rupture due to excessive force. To attenuate large forces and provide damping against large impulsive impact forces, the array of sensors is covered by a layer of compliant material such as rubber or polyimide.

The use of a compliant layer for fingertip contact serves four beneficial functions.

(1)As mentioned previously the compliant layer serves to protect the sensors from damage due to contact forces.

(2) In order to facilitate a stable grasp it is desirable to enhance friction at the contact

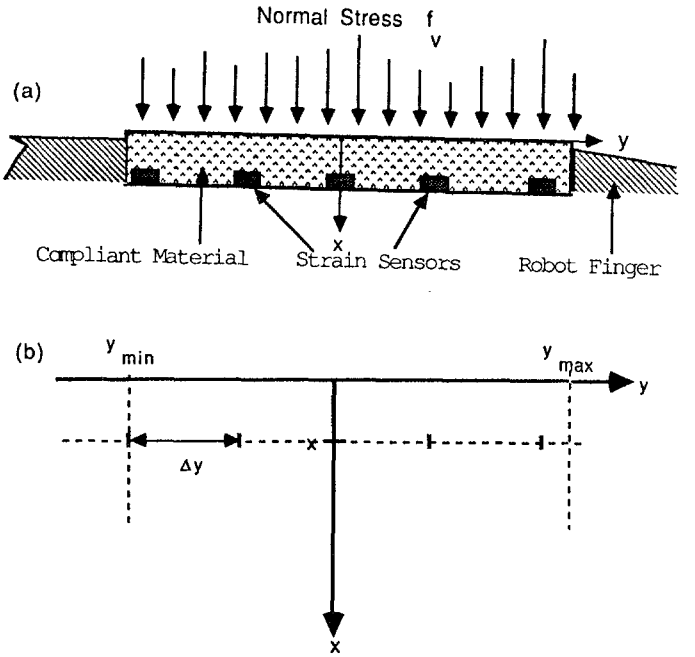


Figure 2: *(a) Stress applied at the boundary of elastic layer with strain sensors beneath surface. (b) Elastic layer extends from y_{min} to y_{max} with sensors at depth x beneath the surface at intervals of Δy.*

surfaces. For given material at the contact surface, this can be achieved by maximizing contact area. A rigid material in contact with an irregular or bumpy surface, contacts the surface at a few discrete points only whereas a compliant material can conform to the surface and thus maximize contact area Since the tactile sensing array provides the desired contact surface on the robot finger, it is beneficial to provide compliant contact atop the array of sensors.

(3)The third function of the compliant layer is due to the resultant 'blurring' of the contact force distribution which causes information about the entire stress distribution to be passed to each individual sensor element. In the absence of such blurring, determination of the applied stress between sample points would not be possible.

(4) Compliant contact is beneficial in establishing well-posedness of the grasp problem (see [3]).

Referring to Figure 2(a), we state a fundamental inverse problem.

Inverse Problem: Given samples of a strain distribution, measured by sensors at a given depth beneath the surface of a compliant layer, the inverse problem of tactile sensing refers to the problem of determining the surface stress distribution which induced the measured strain.

The inverse problem as stated above is formulated for the particular tactile sensor design which we consider here, but is similar to the inverse problems associated with many other tactile sensors as well.

2.2 Modeling the Transduction Process

For sake of simplicity in the current discussion, two assumptions are made. (i) The general three dimensional problem is reduced to a two dimensional setting. i.e. we consider a linear

array of sensors with planar stress applied to the pad. (ii) It is also assumed that the compliant layer is actually a homogeneous, isotropic, linear, elastic half-space.

To understand the inverse problem, a model for the forward transduction process is first developed i.e. a model describing the relationship between stress applied to the compliant layer and the strain induced at a depth x beneath the surface. From the theory of elasticity (see Timoshenko and Goodier [54]) two relationships (for detailed derivations see [55] and [11]) can be derived..

Strain at a depth x beneath the surface of the compliant material due to normal stress at the surface is given by

$$p_x(y) = (Kq_v)(y) = \int_{-\infty}^{\infty} k_x^v(y, y_0) q_v(y_0) dy_0 \tag{1}$$

where, $p_x(\cdot)$ is the strain at depth x, $q_v(\cdot)$ is the surface stress, and $k_x^v(\cdot,\cdot)$ is the convolution kernel relating the two given by,

$$k_x^v(y, y_0) = \frac{2x((1-\nu)^2x^2 - \nu(\nu+1)(y-y_0)^2)}{\pi E(x^2+(y-y_0)^2)^2} \tag{2}$$

where, ν is Poisson's ratio for the material and E is the modulus of elasticity. Strain due to tangential stress applied at the surface is given by an analogous formula, with convolution kernel,

$$k_x^t(y, y_0) = \frac{2(y-y_0)((1-\nu)^2x^2 - \nu(\nu+1)(y-y_0)^2)}{\pi E(x^2+(y-y_0)^2)^2}. \tag{3}$$

For the sake of simplicity we only consider the case of normal stress throughout this paper since the case with tangential stress is obviously analogous[2].

Since measurements of strain are made at a discrete number of points only, equation (1) must be discretized.

Assume that the sensors are distributed uniformly (equal spacing) beneath the surface of the compliant layer. Let $\Delta_p y$ be the distance between points at which strain is sampled. So,

$$\Delta_p y = \frac{y_{max} - y_{min}}{N-1}$$

(see Figure 2), where N is the total number of sensors. Let $\Delta_q y$ be the distance between points where the stress profile is to be reconstructed. Although it is not necessary for $\Delta_p y$ and $\Delta_q y$ to be equal, for the current discussion we let $\Delta_p y = \Delta_q y = \Delta y$. To obtain a discretized version of equation (1), let ϵ_x be the vector of strain samples i.e., $\epsilon_x = (\epsilon_{x_1}, \ldots, \epsilon_{x_N})^T$. Therefore, $\epsilon_{x_i} = p_x(y_{min} + (i-1)\Delta y) \qquad i = 1, \ldots, N$. Similarly, let $f_v = (f_{v_1}, \ldots, f_{v_N})^T$ be the vector obtained from the stress distribution as, $f_{v_i} = q_v(y_{min} + (i-1)\Delta y) \qquad i = 1, \ldots, N$. The convolution kernel $k_x^v(\cdot,\cdot)$ can be discretized to form the matrix $T = \{T_{ij}\}$ by letting $T_{ij} = k_x^v(y_i, y_j)$, where $y_i = y_{min} + (i-1)\Delta y$ for $i = 1, \ldots, N$. Hence the discretization of equation (1) results in,

$$\epsilon_x = T \cdot f_v. \tag{4}$$

The discretized inverse problem is precisely the problem of *determining f_v given ϵ_x and T.*

Returning to the third desirable feature of compliant contact, we observe that if we let T be a non-square matrix then in solving the inverse problem we are attempting to reconstruct

[2]The analogy referred to here does not extend to the choice of regularizer for the case of tangential stress.

the surface stress at points along the surface other than those directly above the sensors. This is only possible since the blurring of signals by the compliant layer results in information about most of the stress distribution at the surface being passed to every sensor. The exceptions to this are due to zeros in the convolution kernel.

It can be shown (see Appendix A) that in the infinite dimensional setting of equation (1), the inverse problem is ill-posed. Since the operator K is a compact operator on an infinite dimensional domain, its inverse does not exist as a bounded operator. Hence Hadamard's third requirement for well-posedness is violated. It can also be shown (see Appendix A) that the the manifestation of this ill-posedness in the discretized problem (4) is occurs in the ill-conditioning of the matrix T. Hence solutions to the inverse problem in the discretized case are sensitive to noise in the data. Ill-conditioning of the matrix T also increases with the dimension of T.

2.3 Regularization as a Technique to Solve the Inverse Problem

There exists a large body of literature devoted to approximating the solutions of ill-posed problems (see e.g. Tikhonov [52] and Tikhonov and Arsenin [53]). One successful technique for solving ill-posed problem is regularization which was introduced by Tikhonov [52] in 1963. Ill-posed problems such as the one considered here are often insufficiently constrained and require the imposition of additional constraints for the solution to be well defined. Regularization is a technique in which the problem is formulated as a variational principle which is then used to impose physical constraints on the solution. A variational principle defines the solution to a problem as the function which minimizes an appropriate cost functional (Poggio and Koch [41]).

Regularization requires the choice of a norm $||\cdot||$ and of a stabilizing functional (typically of the form $||Px||$). The stabilizing functional embodies the physical constraints of the problem and thus must be chosen only after careful analysis of the physical setting in which the problem arises. Constraints such as smoothness and boundedness of solutions may be imposed by appropriate choice of the stabilizer.

The problem of solving $Kx = y$ can be formulated as a variational principle simply by choosing a norm $||\cdot||$ and the finding x which minimizes

$$||Kx - y||$$

[3] To regularize the problem additional constraints are imposed through the stabilizing functional.

Standard regularization theory is composed primarily of three methods (see [41] and [42]).

1. Among all x which satisfies the condition $||Px|| < c$, where c is a constant, find x which minimizes $||Kx - y||$.

2. Among all x which satisfies $||Kx - y|| < \epsilon$,where ϵ is chosen to represent estimated errors, find x which minimizes $||Px||$.

3. Find x which minimizes
$$||Kx - y||^2 + \lambda||Px||^2 \tag{5}$$
where λ is called the *regularization parameter.*

[3] This is easily shown to be equivalent to using the generalized inverse $K^\dagger$ to obtain solutiona; i.e. letting K^* denote the adjoint of K, $x = (K^*K)^{-1}K^*y = K^\dagger y$ whenever the inverse on the right exsits.

In standard regularization theory, the operator P is linear and the norm $||\cdot||$ is derived from an inner product. For such quadratic variational principles, of the form (5), it can be shown that under mild conditions the solution space is convex (which implies the existence, uniqueness and stability of solutions). In this paper we will consider other forms of the stabilizing functional which we will denote by $M(x)$. Hence we will be considering variational principles of the from

$$||Kx - y||^2 + \lambda M(x) \tag{6}$$

The regularization parameter λ controls the degree to which a solution is regularized. Small values of λ compromise the degree of regularization in favor of accurately matching the initial data. Very large values of λ may result in very regular but unrealistic solutions.

2.3.1 Regularizing the Tactile Sensing Problem

To regularize the inverse problem of tactile sensing it is necessary to first identify the generic physical constraints that may be imposed upon the solution. In the case of normal stress applied to the the compliant pad (see Figure 2), it is clear that the unisense nature of the compressive loading on the boundary can be captured by constraining solutions to lie in the positive orthant. To further suppress some of the deleterious effects of sensor noise, the solutions may be constrained to be smooth. Constraining solutions to be smooth may result in inaccurate solutions near physical edges, however, edges may be recovered in a second stage of regularization.

The constraints of nonnegativity and smoothness of the solutions can be embodied in the stabilizing functional by choosing

$$M(x) = \sum_i x_i \log x_i \tag{7}$$

which has the same functional form as Shannon entropy. Hence in the case of the inverse problem of tactile sensing, the problem is to find a vector $f \in \mathbb{R}^N$ which minimizes

$$\|Tf_v - \epsilon_x\|^2 + \lambda \sum_i f_{v_i} \log f_{v_i} \tag{8}$$

where $T \in \mathbb{R}^{N \times N}$ is a finite matrix approximation to the convolution operator, $\epsilon_x \in \mathbb{R}^N$ is the vector of measured strains and $f_v \in \mathbb{R}^N$ is the vector of stress components. (The norm $||\cdot||$ is the standard Euclidean norm on $\mathbb{R}^N$

Exsistence and uniqueness of solutions to the minimization problem are easily verified by noting that the positive orthant in $\mathbb{R}^N$ is a convex set and that equation (8) is a strictly convex function of f_v.

3 Nonlinear Electrical RC Networks

Nonlinear electrical networks have been a topic of active research for many years. Hence there exists a large body of results pertaining to such networks. As elaborated in Section 1, the transition from neural networks to electrical networks is not only trivial, but natural. Mathematical analysis of neural networks is as yet a developing field. It seems natural to look to the available results for nonlinear electrical networks for insight and understanding of the behavior of neural networks.

Some of the earliest work on electrical networks was done by James Clerk Maxwell in 1873 [36] and is concerned with the distribution of currents and voltages in linear resistive

networks. Later work by Tellegen [49], Cherry (1951) [4], and Millar (1951) [37] helped to build a foundation for nonlinear network analysis. In 1964 Brayton and Moser [2] attempted to build a more general theory of nonlinear networks by considering some geometric aspects of such networks. More recently, several researchers have adopted an even more general geometric view of nonlinear networks (see e.g. [35], [9], [34],[33], and [46]. In these latter works, network dynamics are viewed as flows (differential equations) on nontrivial manifolds (nonlinear spaces). It is clear that a great deal of mathematical machinery has been developed for analysis of nonlinear networks. Applications of the same tools and body of results to neural networks should prove useful.

In this section one particular application of electrical network analysis to neural networks is demonstrated. We define an energy function for a dynamical system as a functional which is minimized (globally or locally) as a result of the natural time evolution of the system. We present in this section a theorem which is then applied in Section 4 to determine an energy function for a neural network designed to solve the inverse problem described in Section 2.

3.1 A Theorem on Energy Functions for Nonlinear RC Networks

As in [2], we consider a network composed of branches and nodes with the restriction that a branch connects exactly two nodes. Arbitrarily assigning a direction to the branch currents, we define i_μ as the current flowing from the initial node to the end node of the μth branch in the network. The branch voltage v_μ is defined as the voltage rise measured from the end node to the initial node of the μth branch in the network.

For any network, a complete set of generalized current or voltage coordinates can be chosen. Such a set of variables is complete in the sense that they can be assigned values independently without violation of Kirchoff's laws and that they determine in each branch of the network one of the two variables, branch current or branch voltage. In computing M, the number of defining current coordinates, constant current sources if any are not counted as branches and similarly in computing N, the number of defining voltage coordinates, constant voltage sources are not counted as branches. In the particular case of a RC network, a complete set of variables is obtained by considering the voltages across all independent capacitive branches. Two capacitors in parallel are considered as a single capacitive branch with capacitance equal to the parallel combination of the two. We will denote the complete set of variables for a RC network by $v^* = (v_1, \ldots, v_N)$.

The following theorem (which appears in [40]) identifies an energy function for a class of nonlinear RC networks. For the proof of this theorem see Appendix B.

Theorem 3.1 *Consider a nonlinear RC electrical network for which the following hypotheses hold.*

H1: The voltages across independent capacitive branches, $v^* = (v_1, \ldots, v_N)$ form a complete set of variables for the network.

H2: Let $i_1, \ldots, i_N$ be the currents through the corresponding capacitive branches (with the appropriate reduction of parallel capacitors) satisfying,

$$\frac{\partial i_k}{\partial v_j} = \frac{\partial i_j}{\partial v_k} \qquad j,k = 1,\ldots,N \quad j \neq k$$

Then,

1. *The equilibrium states of the network correspond to the stationary points of the energy function,*

$$P(v^*) = -\sum_{j=1}^{N} \int_0^{v_j} i_j \, dv_j$$

2. *Stable equilibrium states of the network correspond to local minima of $P(v^*)$.*

3. *If in addition to this $P(v^*) \to \infty$ as $||v^*|| \to \infty$, then the network is asymptotically stable i.e. v^* will approach one of the stable equilibrium states of the network (given by the local minima of $P(v^*)$) as $t \to \infty$.* ■

As a corollary to the preceding theorem, we consider the case where we would like to formulate the energy function in terms of an auxillary set of independent variables. For proof of the corollary see [40].

Corollary 3.1 *Assume (H1) and (H2) of Theorem 3.1 hold. Let $u^* = (u_1, \ldots, u_N)$ be the auxillary set of independent variables in which we are interested and let $u_i = g_i(v_i)$, $i = 1, \ldots, N$, where $g_i : \mathbb{R} \to \mathbb{R}$. Then if $g_i(\cdot)$ $i = 1, \ldots, N$ are monotone increasing functions, then the conclusions of Theorem 3.1 hold and the energy function $P(\cdot)$ may be expressed in terms of the variables u^* by simply replacing each v_i by $g_i^{-1}(u_i)$.*

4 An Analog Neural Network Solution To The Inverse Problem

In Section 1 it was noted that, in all but the most rudimentary applications of tactile sensing, real-time processing of tactile sensory data is crucial. In Section 2 it was shown that the inverse problem of tactile sensing may be formulated as a variational principle. Furthermore, the inclusion of a regularizing penalty functional in the variational principle provides immunity to sensor noise and transforms the problem to one that is well-posed. Solutions to the inverse problem (obtained as solutions to a variational principle, require the minimization of the nonlinear cost functional in equation (8). Real-time solutions to such a minimization problem would require fairly powerful digital hardware using standard iterative algorithms. Since it is our objective to obtain fast solutions to the inverse problem as well as limit the hardware complexity to that which can be located in close proximity to individual sensor arrays, analog neural networks provide an attractive alternative.

In this section an analog neural network is described which has been structured so as to solve the regularized inverse problem of tactile sensing. The network described exploits inherent parallelism in the problem since solutions are obtained as the result of asynchronous relaxation of all variables in the network.

We show that the network described, satisfies the hypotheses of Theorem 3.1 and thus an energy function for the network is explicitly given by the theorem. It is shown that the resultant energy function is indeed the variational principle of equation (8).

Computer simulations of the network are used to evaluate the effect of the regularizing parameter λ on the resultant solutions and to evaluate performance of the network in the presence of sensor noise.

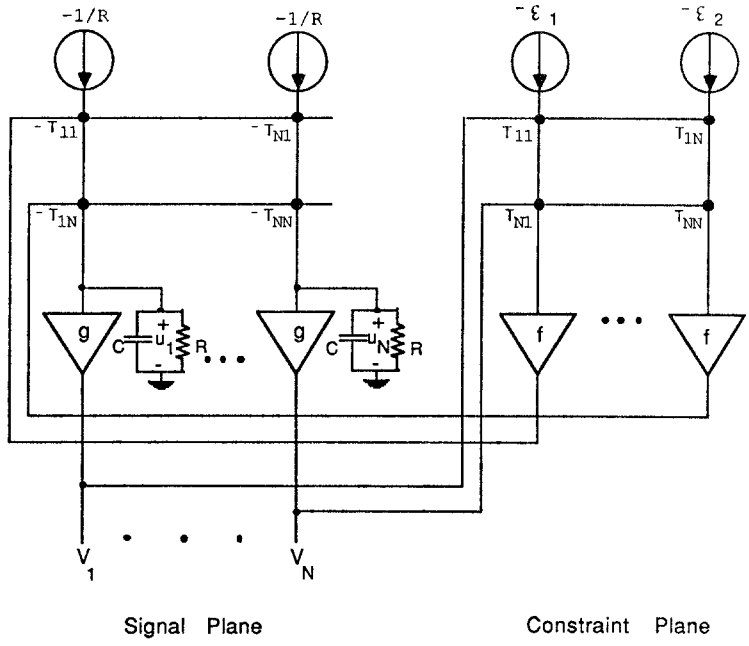


Figure 3: *N-Channel Maximum Entropy Deconvolution Network.*

4.1 The Maximum Entropy Deconvolution Network

Inspired by the work of Tank and Hopfield (see [47]) on neural networks for solving optimization problems, Marrian and Peckerar [28] suggested a neural network for solving deconvolution problems with an entropy-like regularizer. Figure 3 shows a schematic of a N channel deconvolution network.

The network proposed consists of two planes of amplifiers (i) the signal plane and (ii) the constraint plane (see Figure 3). Inputs to the constraint plane ($\epsilon = (\epsilon_1, \ldots, \epsilon_N)^T$) are currents proportional to sample strain measurements obtained from an array of tactile sensors. Outputs of the signal plane ($u = (u_1, \ldots, u_N)^T$) are voltages which, in equilibrium, represent regularized solutions to the inverse problem. The interconnections T_{ij} are conductances (resistors with values $1/T_{ij}$) corresponding to elements of the matrix representation T of the discretized convolution kernel. Amplifiers in the signal plane are exponential i.e. $g(x) = \exp(x)$ and constraint plane amplifiers are linear with gain s ($f(x) = sx$). To intuitively understand the manner in which this deconvolution network operates, dynamical evolution of the network can be viewed as occuring in a series of infinitesimal discrete time steps. If evolution of the network is viewed in this manner, the feedback loops generate a number of 'analog iterations'. Each analog iteration is approximately composed of the following steps:

1. Outputs of the signal plane are convolved with the discretized kernel T in the constraint plane, forming the vector $T \cdot u$.

2. Error in the current estimate of the solution (given by outputs of the signal plane) is evaluated in the constraint plane by subtracting the input strain vector ϵ from the results of the previous step i.e the vector $T \cdot u - \epsilon$ is formed and fed back to the signal plane through the constraint plane amplifiers with gain s.

3. Based upon feedback from the constraint plane, the outputs of the signal plane are updated so as to reduce the error.

4. If outputs of the signal plane have not yet settled repeat (1)–(3)

It remains to be shown that the above series of 'iterations' do indeed converge, i.e. that the network, as an analog electrical circuit, is asymptotically stable. Using the results from Section 3, we now show that the energy function for the network shown in Figure 3 corresponds to the variational principle of equation (8). Thus stable equilibrium states of the network correspond to regularized solutions to the inverse problem. Since minimization of equation (8 also involves maximizing the entropy of the solution, we will refer to the network in Figure 3 as the *maximum entropy deconvolution network* (or MaxEnt network for short).

Stability of the MaxEnt deconvolution network can be established by applying Theorem 3.1 which also determines explicitly an energy function for the network. In the following it is assumed that for the MaxEnt network, any dynamics associated with the constraint plane amplifiers are negligible. This assumption is reasonable since the feedback capacitor associated with the signal plane amplifiers can be chosen so that the response of the signal plane amplifiers is sufficiently slower than those of the constraint plane.

In order to apply Theorem 3.1 it is necessary to first verify that the network satisfies the two hypotheses. From Figure 3 it is clear that the voltages across the signal plane capacitors $v_1, \ldots, v_N$ form a complete set of variables for the network, i.e $v_1, \ldots, v_N$ can be assigned values arbitrarily and that they determine in every branch of the network one of the two variables, branch current or node voltage.

From Figure 3, the current through the capacitor connected to the nth signal plane node is given by,

$$i_n = C\frac{dv_n}{dt} = -\frac{v_n}{R} - \frac{1}{R} - \sum_k t_{kn} f(T_k \cdot u - \epsilon_k). \tag{9}$$

Here $T_k = (t_{k1}, t_{k2}, \ldots, t_{kN})^T$. It is easily verified from (9) that the hypothesis *H2* of Theorem 3.1 is satisfied using $v_1, \ldots, v_N$ as the set of generalized voltage coordinates for the network. Since we are interested in the behavior of the network outputs $u = (u_1, \ldots, u_N)$, we note that u_i are related to v_i by a monotone increasing function $u_i = \exp(v_i)$ $i = 1, \ldots, N$, and apply Corollary 3.1 to write the energy function for the network in terms of the output variables u.

$$P(u) = -\sum_n \int_0^{u_n} i_n du_n \tag{10}$$

The dynamical equations of the network can now be written in the following form:

$$\frac{du}{dt} = -GC^{-1}\frac{\partial P(u)}{\partial u}, \tag{11}$$

where $u = (u_1 \ldots, u_N)$, $v_n = g^{-1}(u_n)$, $G = diag(g'(v_1), \ldots, g'(v_N))$, and $C = diag(C_1, \ldots, C_N)$. Because $g(v) = \exp(v)$, is a monotone increasing function G is always positive definite. Hence from (11), it is clear that the equilibrium states of the network must correspond to stationary points of $P(\cdot)$.

In order to understand the nature of the equilibrium states (if any exist) of the network we must evaluate the integral expression for P.

$$P(u) = \sum_n \int_0^{u_n} \left(\frac{g^{-1}(u_n)}{R} + \frac{1}{R} + \sum_k t_{kn} f(T_k \cdot u - \epsilon_k) \right) dv_n \tag{12}$$

$$= \sum_n \int_0^{u_n} \frac{g^{-1}(u_n)}{R} du_n + \sum_n \int_0^{u_n} \frac{du_n}{R} \tag{13}$$

$$+ \sum_n \int_0^{u_n} \sum_k t_{kn} f(T_k \cdot u - \epsilon_k) du_n. \tag{14}$$

If we let $F(z_k)$ be such that $dF(z_k)/dz_k = f(z_k)$ then,

$$P(u) = \sum_n \int_0^{u_n} \frac{g^{-1}(u_n)}{R} du_n + \sum_n \frac{u_n}{R} + \sum_k F(T_k \cdot u - \epsilon_k). \tag{15}$$

Since the signal plane amplifiers are characterized by $g(u) = \exp(u)$, and the constraint plane amplifiers are characterized by $f(z) = sz$ where s is a constant defining the feedback gain,

$$P(u) = \sum_k \frac{s}{2}(T_k \cdot u - \epsilon_k)^2 + \frac{1}{R}\sum_n u_n \log u_n. \tag{16}$$

It is clear from (16) that $P(u) \to \infty$ as $||u|| \to \infty$. Hence all solutions to (11) approach one of the set of equilibrium states as $t \to \infty$. In Section 2 it was noted that there exists a unique minimum of (16) which correspond to the regularized solution of the inverse problem. Therefore outputs of the network will converge to a regularized solution of the inverse problem.

Equation (16) gives us an explicit form for the energy function for the maximum entropy deconvolution network.

Introduction of other Regularizers As discussed in Section 2, the choice of a regularizing principle must be based upon the physical constraints present in the problem. Also in Section 2 it was shown that the entropy regularizer is appropriate for the recovery of a stress distribution which is normal to the compliant sensing pad. In the case of tangential stress distributions, use of an entropy regularizer would be inappropriate since tangential stress is not unisense in nature. However, it can be seen from equation (15) that the network structure described in the last section is not restricted to use with an entropy regularizer only. It is clear that any regularizer which can be written in the form:

$$\sum_{i-1}^{N} \int_0^{v_i} g^{-1}(v) dv \tag{17}$$

can be introduced into the energy function of the network provided $g(\cdot)$ is a monotone increasing function which can be implemented as the characteristic function of an analog amplifier.

4.2 Simulations

Computer simulations of the maximum entropy deconvolution network were performed in order to better understand behavior of the network in terms of speed of convergence, noise immunity and the effect of the regularizing parameter λ. SIMNON[4], an interactive simulation program for nonlinear dynamical systems was used to simulate the network.

For the purpose of simulation, the stress distribution at the surface of the compliant layer was assumed to be the result of applying pressure normal to the surface using a cylindrical object (see Figure 4). The resulting normal stress distribution due to such a cylindrical indentor can be written as (see [6]),

$$f_v(y) = \begin{cases} \frac{p}{\pi a^2}\sqrt{a^2 - y^2} & \text{if } y \in [-a, a] \\ 0 & \text{elsewhere} \end{cases} \tag{18}$$

[4]SIMNON was provided to us by Professor Astrom from the Lund Institute of Technology.

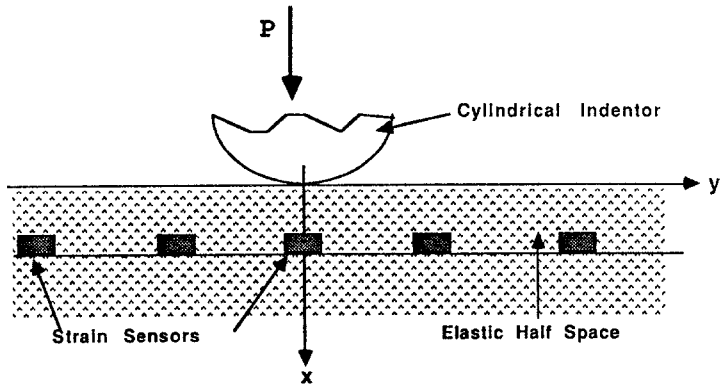


Figure 4: *Application of stress at boundary of elastic half-space using a cylindrical object*

where p is the force per unit length and a is the halfwidth of the contact region. Simulations were performed using $p = 3$ and $a = 1$. The resulting stress distribution is then convolved with the convolution kernel k_x relating normal stress to strain which is given by,

$$k_x^v(y - y_0) = \frac{3}{2\pi E} \frac{x(x^2 - (y - y_0)^2)}{(x^2 + (y - y_0)^2)^2}, \tag{19}$$

where the modulus of elasticity is E and Poisson's ratio for the material is assumed to be 0.5. ($E = 1$ and $x = 1$ were used for ther purpose of simulation.) The resulting 'strain' is then provided as input to the network which then attempts to reconstruct the surface stress distribution.

Accurate asessments of convergence time could not easily be made using digital computer simulations. If the signal plane capacitors were assigned values so as to reduce convergence time (by decreasing the RC time constant) numerical instability resulted. By decreasing the integration time step the problems with instability could be avoided at the cost of tremendous increases in the time necessary to perform a simulation; so it was merely observed at this stage that the network does converge.

To evaluate performance of the network in the presence of noise, Gaussian white noise with variance σ^2 was added to the strain from which surface stress was to be determined. Figure 5 shows the reconstruction of surface stress obtained by the network from noisy data ($\sigma^2 = 0.1$) with $\lambda = 0$ i.e. without any regularization. It can be seen that the reconstruction is very poor due to multiple peaks and negative solutions. However, in comparison to the reconstruction obtained under identical conditions using the Discrete Fourier Transform (see Figure 6) the network solution is markedly superior.

As the regularizing parameter λ is varied (see Figures 7–8), varying degrees of positivity and smoothness are imposed upon the solution. For $\lambda = 0.1$ (Figure 7)it is clear that although the solution has been constrained to the positive orthant, the degree of regularization is insufficient for the given noise conditions. In Figure 8 ($\lambda = 100$) the solution has been over-regularized since solutions which should have been close to zero have been pushed away from zero and the peak of the distribution has been greatly suppressed. Figure 9 shows reconstruction obtained using $\lambda = 10$ which is the 'best' of the three shown.

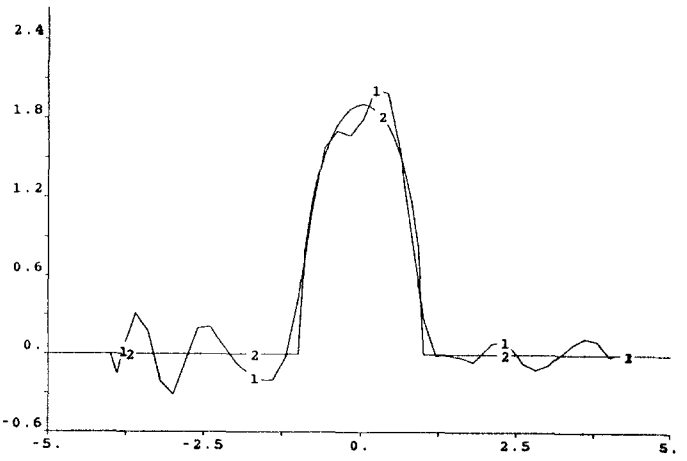


Figure 5: *Network reconstruction of surface stress from noisy strain data ($\sigma^2 = 0.1$) without regularization ($\lambda = 0$) (1)Reconstruction, (2)Designed surface stress.*

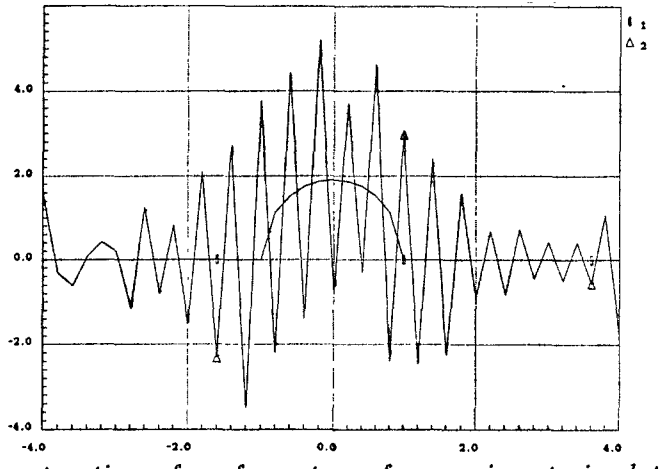


Figure 6: *Reconstruction of surface stress from noisy strain data using the DFT approach (1)Designed surface stress, (2)Reconstruction.*

Clearly, for given noise conditions, there exists a value of the regularizing parameter λ for which the reconstruction is optimal in some sense. Since the objective is to match the reconstruction to the original stress distribution as closely as possible, it is reasonable to choose a value of λ for which the discrepancy between the original stress distribution and the network reconstruction is minimized. One measure of the discrepancy between the known solution and the network solution is the mean squared error which can be written as,

$$MSE = \frac{1}{N}||f_v^0 - f_v||^2 \tag{20}$$

where $f_v^0 \in \mathbb{R}^N$ is the known solution to the inverse problem and $f_v \in \mathbb{R}^N$ is the solution obtained by the network. Such an approach to the problem of choosing λ suggests the use of design tools such as CONSOLE (see [9]) which is a computer aided design tool for parametric optimization of dynamical systems[5].

[5]CONSOLE was used to optimize parameters in design of the breadboard prototype network which was constructed

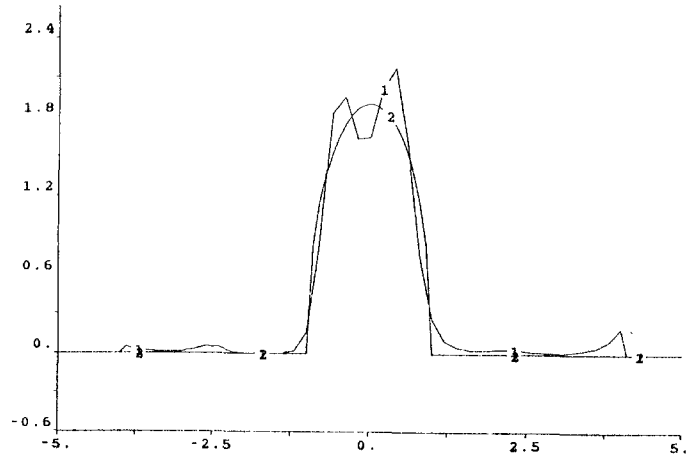


Figure 7: *Network reconstruction of surface stress from noisy strain data for* $\lambda = 0.1$ *(1)Reconstruction, (2)Designed surface stress.*

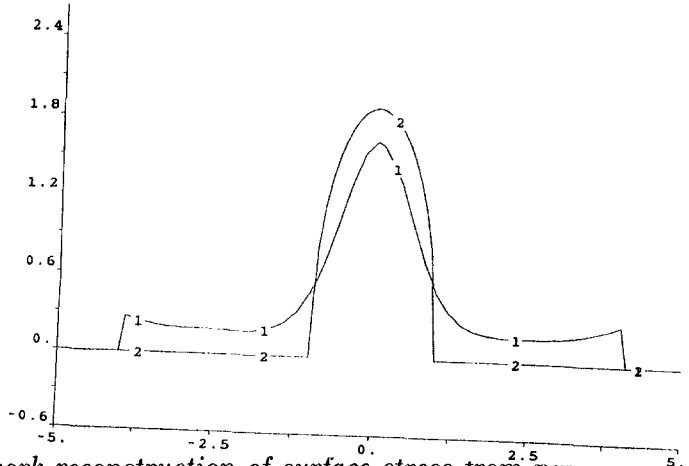


Figure 8: *Network reconstruction of surface stress from noisy strain data for* $\lambda = 100$ *(1)Reconstruction, (2)Designed surface stress.*

4.2.1 Breadboard Prototype of Deconvolution Network

As discussed in earlier, accurate assesments of convergence time for the network are not easily made using digital computer simulations. Also, in the analysis of the deconvolution network , it was assumed that any dynamics associated with the constraint plane could be ignored provided that the signal plane amplifiers are sufficiently slower in response. In practical implementations of such a network, it is necessary to understand what effects delays in the constraint plane response may have upon the network. It is ultimately the constraint plane dynamics which limit the speed of convergence which is achievable. A formal treatment of this subject is to be found in Marcus and Westervelt [26]. A prototype breadboard model of the deconvolution network was constructed using 'off-the-shelf' operational amplifiers, resistors and capacitors. The network was constructed with seven signal plane nodes and seven constraint plane nodes. Since the purpose of constructing the breadboard prototype was to estimate the speed achievable by such a network, exponential amplifiers in the signal plane were replaced by unity gain linear amplifiers to simplify the circuit [6]. Replacing the exponential amplifiers by linear amplifiers

[6] A second breadboard prototype was also constructed which contained the exponential amplifiers, but was used to solve a different problem (see [27]).

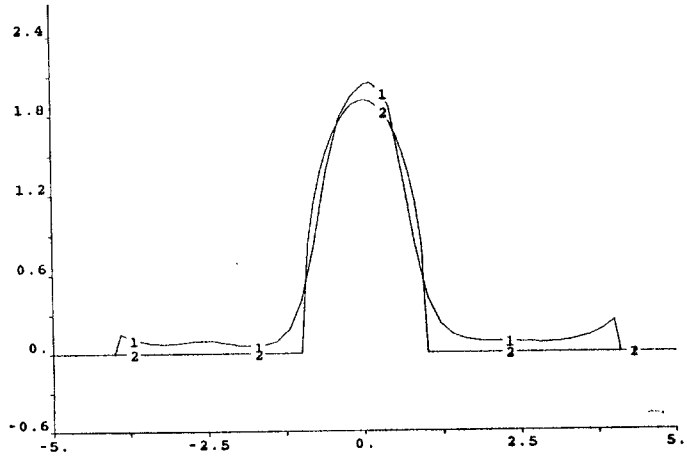


Figure 9: *Network reconstruction of surface stress from noisy strain data for* $\lambda = 10$ *(1)Reconstruction, (2)Designed surface stress.*

results in the entropy regularizer being replaced by a regularizer of the form,

$$\lambda M(v) = \frac{1}{R}\sum_i v_i^2. \tag{21}$$

The interconnection matrix $[T_{ij}]$ was chosen as a seven point discretization of the elastic kernel $k_x^v(\cdot)$ in equation (19) and implemented using resistors with values $R_{ij} = 1/T_{ij}$. Thus if the voltage output v_k of node k is connected to the input of node j through a resistor R_{kj} then current input to node j due to node k is given by Ohm's Law as,

$$i_{jk} = \frac{v_k}{R_{kj}} \tag{22}$$

which is as desired.

Inputs to the network (currents injected into the constraint plane) were chosen to represent samples of the strain distribution due to the compressive loading profile used for the simulations.

The rise time of the constraint plane amplifiers was measured to be approximately 1 μsec. Actual response time of the constraint plane would be longer than this since the parallel combination of all resistors connected to the input of any node contribute to the RC time constant. It was observed that for choices of the signal plane capacitors C for which the rise time of the outputs of the network would be below 10 μsec, the outputs would oscillate i.e. the network was unstable. For C=10 pF the rise time of the outputs of the network was measured to be 10 μsec (see Figure 10). It is clear that the use of faster operational amplifiers would result in an increase in achievable speed since this would decrease the constraint plane response time and thereby permit a decrease in the time constant of the signal plane.

Settling time and overshoot of the outputs of the network are controlled by the gain of the constraint plane nodes. CONSOLE (see [9]) was used to choose a value for the gain so as to minimize overshoot and settling time.

4.2.2 Integrated Circuit Prototype

A prototype analog integrated circuit implementation of the deconvolution network described here has been fabricated, but remains to be tested. A hierarchical design philosophy is practiced in this initial implementation. The deconvolution network may be thought of as composed

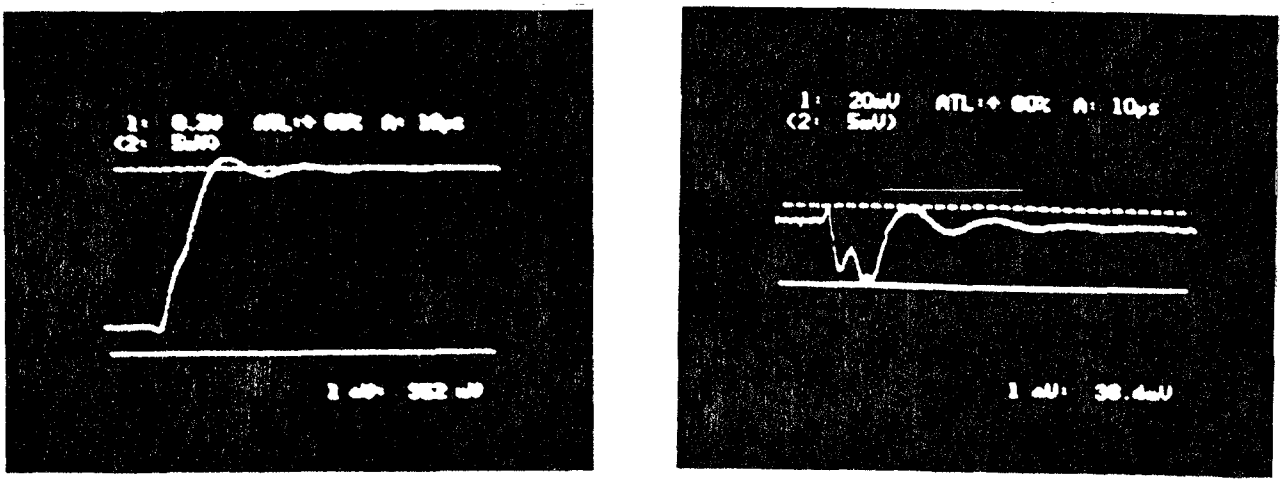


Figure 10: *Oscilloscope trace showing time evolution of a single output of the signal plane for the 7-channel breadboard prototype deconvolution network.*

of two sections: (i) Active components of the network including signal and constraint plane amplifiers and (ii) The functionally passive[7] resistive interconnection matrix. These two sections may also be thought of in the following manner. Once the size of the deconvolution network (number of inputs and outputs) has been decided, the amplifiers of the network are determined. However, the resistive matrix may be a variable entity. For instance, given two different elastic materials (or even two different thicknesses of a given elastic material), the convolution kernel $k_x^v(\cdot)$ and hence its discretization $[T_{ij}]$ are in general different. Thus the network may be thought of as being composed of a fixed part and a variable part.

If fixed resistors are to be used to implement the interconnect matrix, then some provision should be made to change this matrix without having to refabricate the rest of the network. In order to provide some flexibility in the choice of the interconnect matrix and to permit the use of two different fabrication technologies , the deconvolution network was fabricated as two separate integrated circuits.

The amplifier chip (shown in Figure 11, is designed to serve as the 'motherboard' for the network on top of which the resistive connection matrix chip (see Figure 12). Connections between the two chips are made by local wire bonds between bonding pads provided for this purpose on both chips. This approach also facilitates experimentation with different types of connection matrices such as those with programmable connections.

5 Conclusions

In this paper, we considered the inverse problem of recovering stress distributions over regions of contact from samples of strain provided by an array of tactile sensors. In the case where stress is applied to the surface of a compliant material and strain is measured at a fixed depth beneath the surface, the inverse problem was shown to be a problem of deconvolution. It was shown that the technique of regularization could be used to introduce *a priori* knowledge into the problem in order to obtain solutions. The constraints of non-negativity and smoothness were imposed by choosing an entropy regularizer for the recovery of normal surface stress. Solutions to the inverse problem could then be obtained by minimization of a cost functional. We demonstrated that under certain hypotheses, energy functions may be explicitly determined for nonlinear RC

[7]The term 'functionally passive' here is used to describe the fact that in some situations it is desirable to use active circuit components configured to look like a passive resistor from an input/output standpoint.

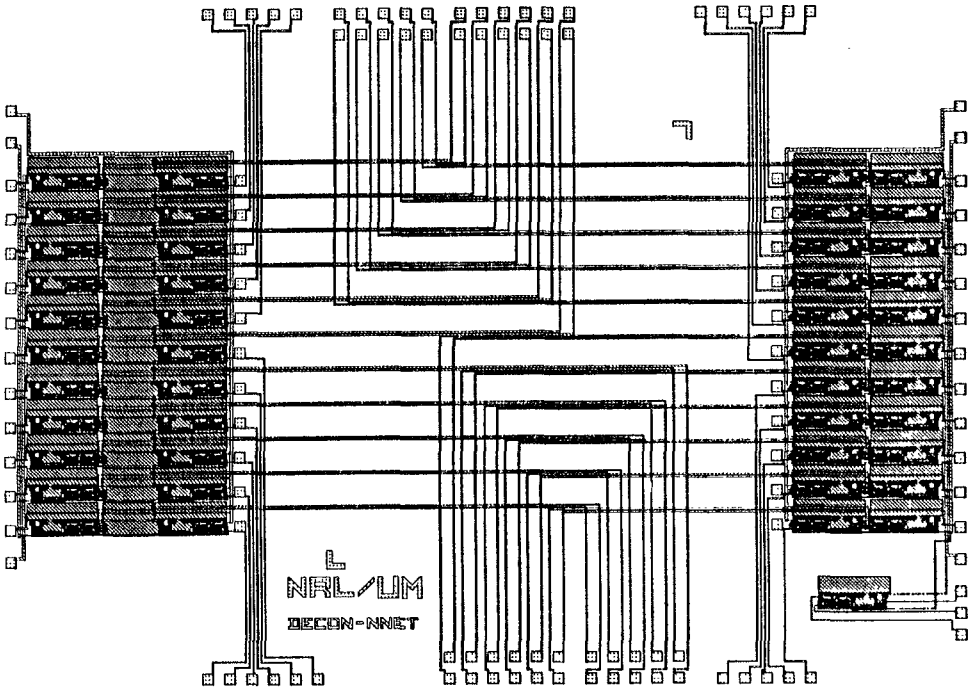


Figure 11: *Layout of integrated circuit chip containing all amplifiers for an eleven channel deconvolution network*

electrical networks. Stable equilibria of the network were shown to correspond to local minima of the energy function and conditions for stability of the network were determined.

An analog neural network for regularized solution of the inverse problem was proposed. Using the results of Section 3, it was shown that the energy function of the proposed network corresponds to the variational principle formulated for solution of the inverse problem of tactile sensing. Stability of the network, in terms of electrical circuit analysis, was guaranteed by monotonicity of the characteristics of the signal plane amplifiers. It was also determined that any regularizer which could be written in the form of equation (17) and satisfied the monotonicity requirements for the signal plane amplifier characteristic g, could be incorporated into the energy function for the network. Computer simulations demonstrated the ability of the deconvolution network to accurately recover normal surface stress even when the sensor outputs were severely corrupted by noise.

A breadboard prototype of the deconvolution network was used to demonstrate the computational speed achievable by such a hardware implementation. Convergence time for the breadboard prototype was measured to be approximately 10μsec. An integrated circuit implementation of the proposed deconvolution network was undertaken. An hierarchical approach was taken in the integrated circuit implementation to provide flexibility in the design of an interconnection matrix.

6 Discussion

Most early vision problems are ill-posed and regularization has successfully been applied in solving many of them (see e.g. [41] and [40]). However, regularization as a technique for solving

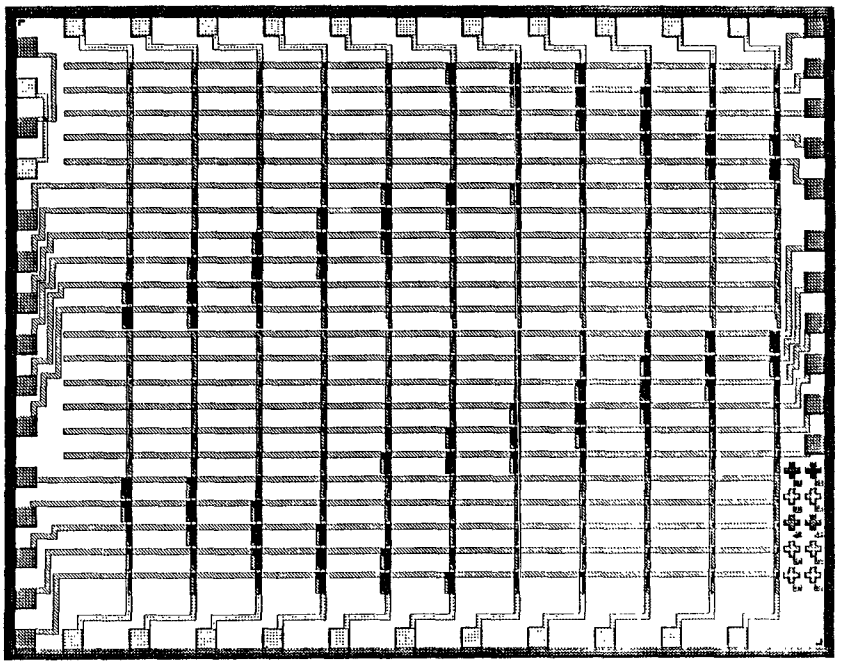


Figure 12: *Layout of resistive interconnection matrix chip.*

ill-posed problems has also demonstrated limitations. Among the limitations of standard regularization theory, is its inability to effectively cope with discontinuities [31]. If the operators K and P in equation (5) are linear (as in standard regularization theory), the solution space is essentially restricted to generalized splines. Hence in some cases the resultant solutions will be overly smooth and cannot be trusted at discontinuities such as edges. It has been suggested [50] that after standard regularization, locations where the resultant solution originates a 'large' error in the second (regularizing) term of equation (5) can be identified as locations of discontinuities. A second second stage of regularization can then be performed, using the locations of discontinuities as boundary conditions. This approach requires the choice of a threshold for error in the regularizing term in order to identify discontinuities. Furthermore the task of locating discontinuities is hindered by the smoothing due to the first stage of regularization. Marroquin in [29] proposes a nonquadratic stabilizer which preserves discontinuities in reconstructions of surfaces from depth data by embedding prior knowledge about the geometry of discontinuities.

Another approach to overcome some of the limitations of standard regularization theory, is based on Bayesian estimation and Markov random field models of the image (tactile or visual). This approach (see [31] uses prior knowledge represented in terms of appropriate probability distributions instead of directly restricting the solution space. It can be shown [30] (see also [31] and [41]), that maximizing the *a posteriori* probability, is equivalent to minimizing an expression of the form (5). However, in this case the functional to be minimized is not in general quadratic as a whole.

It is clear that the class of problems for which quadratic variational principles are sufficient, is limited. For every quadratic variational principle, it can be shown that there exists a corresponding linear analog electrical network consisting of resistors, voltage sources, and current sources, which has the same solutions. This fact is used in [40] to synthesize analog resistive networks to solve problems in early vision. In [22], the approach taken to solving a nonquadratic variational principle employs a hybrid analog-digital network which at each iteration (on the digital time scale) solves a quadratic variational principle (in analog). In general, nonquadratic variational principles may posses numerous local minima in addition to

a global minimum (which may or may not exist). The deterministic gradient descent approach taken in [23] demonstrated an ability to perform well (qualitatively) in comparison to statistical annealing which converges to the global solution with probability one if appropriately applied. However, convergence to the global solution cannot be guaranteed and the hybrid analog-digital nature of the network introduces additional hardware complexity.

In this paper, it was shown that a strictly analog network can be structered so as to solve a nonquadratic variational principle. Convergence to the global solution, in this case, is guaranteed since the variational principle of (8) is strictly convex. It was also shown that there exists a class of nonquadratic variational principles (see equation (17)) which can be solved by similar networks by choosing appropriate characteristics $g(\cdot)$ for the signal plane amplifiers. In the case of this larger class of nonquadratic variational principles, convergence to global solutions cannot in general be guaranteed since multiple minima may exist. Techniques such as adding noise to the network can be used to aid in escaping from local minima in an attempt to find the global solution. Two interesting questions arise in this context: (i)For any nonquadratic variational principle, under what conditions does there exist a (possibly nonlinear and active) analog network with the same solution? and (ii)How can optimality of the solutions be guaranteed?

So far the discussion in this paper has been confined primarily to the problems of early taction. The highest level of the tactile sensing hierarchy, which has been ignored so far, is crucial to the usefulness of any tactile sensing system. A higher level description of the tactile environment is the next step beyond the low-level description provided by the processes of early taction. For instance, although a low-level description of a grasped object may be sufficient to secure the object in a stable grasp, it is not adequate to directly identify the object. In this aspect of tactile sensing as well, neural networks may provide a solution. Adaptive neural networks have demonstrated a remarkable ability to 'learn' complex representations and successfully classify patterns based on these representations. Among other applications of such neural network classifiers are handwritten character recognition [14], identification of faces [1], and classification of superposed radar return signals [50]. Successful VLSI implementation of an adaptive neural network classifier for recognition of grasped objects may further reduce the computational load of the central processor.

It was shown by construction of a breadboard circuit that analog hardware implementation of the proposed network leads to convergence times in the order of 10μsec. In order to compare this with digital computation, we note that simple inversion of an $n \times n$ symmetric Toplitz matrix, is of computational complexity $O(n^2)$. For the purpose of a biased comparison (biased in favor of digital computation), we can ignore the regularizer and assume that it takes exactly n operations to invert the matrix A. Then for a modest array of 25 tactile sensors, it would be necessary to perform,

$$25^2 \frac{operations}{solution} \times 10^5 \frac{solutions}{second} = 62.5 \frac{Million\ \ operations}{second}$$

in order to keep up with the processing speed of the analog network. This is clearly not possible for local digital computation. Also as the size of the sensor array increases, the processing time required for digital computation increases quadraticaly. It was noted in that in the case of the analog network, convergence time actually decreases as the size of the problem increases. In [18] it was established through survey that processing times of approximately 1-2ms are sufficiently fast for most tactile sensing applications. Additional available time resulting from the use of analog network processors could be utilized to perform other tasks such as predicting slippage based upon prior and current processing results.

It is clear that there remain a great many unaddressed and unsolved problems in the area of tactile sensing. Tactile sensing has not received the attention of researchers to the same extent as vision has. As in vision, there is a need in tactile sensing to identify and formalize the problems involved and then devise sensible solutions to these problems. We have suggested in this paper that analog neural networks may provide a solution to some of the low-level processing aspects of tactile sensing and possibly even some of the higher level tasks such as object recognition. However, we have only addressed a small portion of a large and complex problem.

7 Acknowledgements

This research was supported in part by the National Science Foundation's Engineering Research Centers Program: NSFD CDR 8803012, the Air Force Office of Scientific Research under contract AFOSR-88-0204 and by the Naval Research Laboratory.

References

[1] Yaser S. Abu-Mostafa and Demetri Psaltis. Optical neural computers. *Scientific American*, 88–95, March 1987.

[2] R.K. Brayton and J.K. Moser. A theory of nonlinear networks i and ii. *Quarterly Of Applied Mathematics*, 22:1–33,81–184, 1964.

[3] R. W. Brockett. Stability and control of grasping. Presented at IEEE conference on Robotics and Automation, 1986.

[4] C. Cherry. Some general theorems for nonlinear systems possesing reactance. *Phil. Mag.*, 42:1161, 1951.

[5] James J. Clark. A magnetic field based compliance matching sensor for high resolution, high compliance tactile sensing. In *Proceedings IEEE Inter. Conf. On Robotics and Automation, Philadelphia, Pa.*, pages 772–777, April 1988.

[6] H.D. Conway and et. al. Normal and shearing contact stress in indented strips and slabs. *Intl. J. Eng. Sci.*, 4:343–359, 1966.

[7] P. Dario and G. Butazzo. An anthropomorphic robot finger for investigating artificial tactile perception. *Intl. J. of Robotics Research*, 6(3), Fall 1987.

[8] P. Dario, D. DeRossi, C. Domenici, and R. Francesconi. Ferroelectric polymer tactile sensors with anthropomorphic features. In *Proceedings Intl. Conf. on Robotics, Atlanta, Ga.*, pages 332–339, March 13-15 1984.

[9] C.A. Desoer and F.F. Wu. Trajectories of nonlinear rlc networks: a geometric approach. *IEEE Transactions on Circuit Theory*, CT-19(6), November 1972.

[10] M.K.H Fan, J. Koninckx, L.S. Wang, and A.L. Tits. *CONSOLE:A CAD tandem for Optimization-Based Design Interacting with Arbitrary Simulators.* Technical Report, University of Maryland, Systems Research Center, 1987.

[11] R.S. Fearing and J.M. Hollerbach. Basic solid mechanics for tactile sensing. *Intl. J. of Robotics Research*, 4(3), Fall 1985.

[12] Israel Gohberg and Seymour Goldberg. *Basic Operator Theory.* Birkhauser, Boston, MA., 1980.

[13] G. Gordon Ed. *Active Touch.* Oxford: Pergamon Press, 1978.

[14] Hans P. Graf, Lawrence D. Jackel, and Wayne E. Hubbard. Vlsi implementation of a neural network model. *IEEE Computer Magazine*, 41–49, March 1988.

[15] A.R. Grahn and L. Astle. Robotic ultrasonic force sensor arrays. Preprint. Bonneville Scientific, Salt Lake City, Utah.

[16] E.A. Guillemin. *Theory of Linear Physical Systems.* John Wiley and Sons, Inc., New York, 1963.

[17] J. Hadamard. *Lectures of the Cauchy Problem in Linear Partial Differential Equations.* Yale University Press, New Haven, Conn., 1923.

[18] L.D. Harmon. Automated tactile sensing. *Intl. J. Robotics Research*, 1(2):3–32, 1982.

[19] L.D. Harmon. *Touch Sensing Technology: A Review.* Technical Report, Society of Manufacturing Engineers, One SME Drive, P.O. Box 930, Dearborn, Mi. 48128, 1980. Technical Report MSR80-03.

[20] Morris W. Hirsch. Convergence in neural nets. In *Proceedings IEEE International Conference On Neural Networks, San Diego, Ca.*, June 1987.

[21] J.J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci., U.S.A.*, 79:2554–2558, April 1982.

[22] J.J. Hopfield. Neurons with graded responses have collective computational properties like those of two-state neurons. *Proc. Natl. Acad. Sci., U.S.A.*, 81:3088–3092, May 1984.

[23] James Hutchinson, Christof Koch, Jin Luo, and Carver A. Mead. Computing motion using analog and binary resistive networks. *IEEE Computer Magazine*, 52–63, March 1988.

[24] L.D. Jackel, R.E. Howard, H.P. Graf, B. Straughn, and J.S. Denker. Artificial neural networks for computing. *J. Vac. Sci. Technol. B*, 4(1):61–63, January/February 1986.

[25] F.J. Kub, I.A. Mack, K.K. Moon, C.T. Yao, and J.A. Modolo. Programmable analog synapses for microelectronic neural networks using a hybrid digital-analog approach. In *Proceedings IEEE International Conference On Neural Networks, San Diego, Ca.*, July 1988.

[26] R. Lippmann. An introduction to computing with neural nets. *IEEE ASSP Magazine*, 4–22, April 1987.

[27] C.M. Marcus and R.M. Westervelt. Stability of analog neural networks with delay. In Press. Physical Review A.

[28] C. R. K. Marrian, M. C. Peckerar, I. Mack, and Y. C. Pati. Electronic 'neural' nets for solving ill-posed problems with an entropy regularizer. In J. Skilling, editor, *Maximum Entropy and Bayesian Methods*, pages 371–376, Kluwer Academic Publishers, 1989.

[29] C.R.K. Marrian and M.C. Peckerar. Electronic neural net algorithm for maximum entropy deconvolution. In *Proceedings IEEE First Annual Inter. Conf. On Neural Networks, San Diego, Ca.*, June 1987.

[30] J. Marroquin. Memo 792, MIT, Artificial Intelligence Lab, 1984.

[31] J. Marroquin. *Probabilistic Solution Of Inverse Problems.* Technical Report 860, MIT, Artificial Intelligence Lab, 1985.

[32] J. Marroquin, S. Mitter, and T. Poggio. Probabilistic solution of ill-posed problems in computational vision. *Journal of the American Statistical Association*, 82(397):, March 1987.

[33] T. Matsumoto. Eventually passive nonlinear networks. *IEEE Transactions on Circuits and Systems*, CAS-24(5), May 1977.

[34] T. Matsumoto. On dynamics of electrical networks. *J. of Differential Equations*, 21(1), May 1976.

[35] T. Matsumoto. On several geometic aspects of nonlinear networks. *Journal of the Franklin Institute*, 301(1 and 2), January-February 1976.

[36] James Clerk Maxwell. A treatise on electricity and magnetism. vol. I 3rd Ed., pp.407.

[37] W. Millar. Some general theorems for nonlinear systems possessing resistance. *Phil. Mag.*, 42:1150, 1951.

[38] D.H. Mott, M.H. Lee, and H.R. Nicholls. An experimental very high resolution tactile sensor array. In *Proceedings 4th International Conference on Robot Vision and Sensory Control, London*, 1984.

[39] Y. C. Pati, D. Friedman, P. S. Krishnaprasad, C. T. Yao, M. Peckerar, R. Yang, and C. R. K. Marrian. Neural networks for tactile perception. In *Proceedings IEEE Inter. Conf. On Robotics and Automation, Philadelphia, Pa.*, pages 134–139, April 1988.

[40] Yagyensh C. Pati. *Neural Networks for Low-Level Processing of Tactile Sensory Data.* Master's thesis, University Of Maryland., College Park, MD., 1988. Systems Research Center Technical Report No. SRC-TR-89-8.

[41] T. Poggio and C. Koch. Ill-posed problems in early vision: from computational theory to analogue networks. *Proc. R. Soc. London*, B 226:303–323, 1985.

[42] T. Poggio, V. Torre, and C. Koch. Computational vision and regularization theory. *Nature*, 317(6035):314–319.

[43] M.H. Raibert. An all digital VLSI tactile array sensor. In *Proceedings Intl. Conf. on Robotics, Atlanta Ga.*, pages 314–319, March 13-15 1984.

[44] W. Rudin. *Real and Complex Analysis.* McGraw Hill, 1986.

[45] D.B. Schwartz, R.E. Howard, J.S. Denker, R.W. Epworth, H.P. Graf, W. Hubbard, L.D. Jackel, B.S. Straughn, and D.M. Tennant. Dynamics of microfabricated electronic neural networks. *Appl. Phys. Lett*, 50(16,20), April 1987.

[46] S. Smale. On mathematical foundations of electrical circuit theory. *Journal of Differential Geometry*, 7:193–210, 1972.

[47] D.W. Tank and J.J. Hopfield. Neural computation of decisions in optimization problems. *Biological Cybernetics*, 52:141–152, 1985.

[48] D.W. Tank and J.J. Hopfield. Simple 'neural' optimization networks: an a/d converter, signal decision circuit and a linear programming circuit. *IEEE Transactions on Circuits and Systems*, CAS-33(5):533–541, May 1986.

[49] B.D.H. Tellegen. A general network theorem with applications. *Phillips Research Reports*, 7, 1952.

[50] Anthony Teolis, Yagyensh C. Pati, Martin C. Peckerar, and Shihab Shamma. *Cascaded Neural-Analog Networks For Real Time Decomposition of Superposed Radar Return Signals in the Presence of Noise.* Technical Report SRC TR 89-33, University of Maryland, Systems Research Center, 1989.

[51] D. Terzopoulus. *Multiresolution Computation of Visible-Surface Representation.* PhD thesis, Masachusettes Institute of Technology, Cambridge, MA., 1984. Ph.D. Thesis, Department of Electrical Engineering and Computational Science.

[52] A.N. Tikhonov. *Sov. Math. Dokl*, 4:1035–1038, 1963.

[53] A.N. Tikhonov and Arsenin V.Y. *Solutions of Ill Posed Problems.* Winston Press, Washington, D.C., 1977.

[54] S. Timoshenko and J.N. Goodier. *Theory of Elasticity.* McGraw Hill, New York, 1951.

[55] Rui Yang. *Tactile perception for Multi-fingered Hands.* Master's thesis, Univ. Of Md., College Park, MD., 1987. Systems Research Center Tech. Report No. SRC-TR-87-126.

[56] C. T. Yao, M. C. Peckerar, J. Wasilik, C. Amazeen, and S. Bishop. A novel three dimensional microstructure fabrication technique for a triaxial sensor. Preprint. Naval Research Laboratory, Code 6804, Washington D.C., 1987.

Appendix

A Ill-Posedness of the Inverse Problem of Tactile Sensing

The general form of integral equations of the first kind is given by,

$$g(t) = \int_a^b k(t,s)f(s)\,ds \qquad c \leq t \leq d. \tag{23}$$

Equation (23) may be rewritten in operator notation as,

$$g(t) = (Kf)(t) \tag{24}$$

where K is the integral operator with kernel function k. In the particular case of the tactile sensing problem, K is a convolution operator with the convolution kernel k. Since $g(\cdot)$ and $f(\cdot)$ in this case represent strain and stress respectively and are therefore signals with finite energy, we consider the case where $g, f \in L_2([a,b])$ with norm defined by,

$$||h||_2 = \left(\int_a^b |h(x)|^2\,dx\right)^{1/2} \tag{25}$$

and $a = y_{min}, b = y_{max}$. Thus we are interested in the case when, $k_x(y,y_0)$[8]$\in L_2([a,b]\times[a,b])$ and

$$p_x(y) = (Kq_v)(y) = \int_a^b k_x(y,y_0)q_v(y_0)dy_0. \tag{26}$$

To demonstrate the ill-posed nature of the inverse problem, we would like to do the following,

1. Show that in the infinite dimensional setting of equation (26) the inverse of the integral operator K is unbounded.
2. Verify that the finite matrix representation of K in equation (4) (obtained by discretization of the kernel function) is indeed a justifiable approximation of the operator K.
3. Show that the finite dimensional manifestation of the unboundedness of the inverse of K occurs in the ill-conditioning of the finite matrix representation of K.

In order to clarify latter discussion we present some definitions, notation, and theorems. Proofs of the theorems are to found in Gohberg and Goldberg [12] and/or Rudin[44].

Notation:

H, H_i $\quad i = 1,2,\ldots$	Hilbert Spaces
$L(X)$	Bounded linear operators on X
$L(X,Y)$	Bounded linear operators from X to Y
$Sp\{\phi_1,\phi_2,\ldots\}$	Span of $\{\phi_1,\phi_2,\ldots\}$
$Im(A)$	Image of A
A^*	Adjoint of the operator A

Definition A.1 *An operator $A \in L(H_1,H_2)$ is compact if for each sequence $\{x_n\}$ in H_1, such that $||x_n|| = 1$, the sequence $\{Ax_n\}$ has a subsequence which converges in H_2.*

[8] We will use $k_x(y,y_0)$ to denote $k_x^v(y,y_0)$ unless otherwise indicated

Lemma A.1 *If $A \in L(H_1, H_2)$ is of finite rank then A is compact.* ■

Theorem A.1 *Let $\{A_n\}$ be a sequence of compact operators in $L(H_1, H_2)$ such that $||A_n - A|| \to 0$ as $n \to \infty$, where $A \in L(H_1, H_2)$. Then A is compact.* ■

To show that the integral operator K is bounded, we note that since $k \in L_2([a,b] \times [a,b])$,

$$\int_a^b \int_a^b |k(t,s)|^2 \, ds \, dt \; < \; \infty. \tag{27}$$

Now,

$$(Kf)(t) = \int_a^b k(t,s) f(s) \, ds. \tag{28}$$

So by Schwarz's inequality,

$$\int_a^b |k(t,s)f(s)| \, ds \leq (\int_a^b |k(t,s)|^2 \, ds)^{1/2} (\int_a^b |f(s)|^2 \, ds)^{1/2}. \tag{29}$$

Therefore,

$$||Kf||^2 \leq \int_a^b (\int_a^b |k(t,s)f(s)| \, ds)^2 \leq ||f||^2 \int_a^b \int_a^b |k(t,s)|^2 \, ds \, dt. \tag{30}$$

So,

$$||K||^2 \leq \int_a^b \int_a^b |k(t,s)|^2 \, ds \, dt \; < \infty. \tag{31}$$

Hence the operator K is bounded. K is clearly linear as well and so $K \in L(H)$ where $H = L_2([a,b])$. Since from equation (2) we see that $k(t,s) = k(s,t)$, the operator K is also self adjoint.

To show that K is a compact operator we construct a sequence of operators of finite rank and then use Theorem A.1.

Let $\phi_1, \phi_2, \ldots$ be an orthonormal basis for $L_2([a,b])$. Then,

$$\Phi_{ij}(t,s) = \phi_i(t)\bar{\phi}_j(s) \qquad i,j = 1,2,\ldots$$

is an orthonormal basis for $L_2[a,b] \times [a,b])$ (see Gohberg and Goldberg [12]). Therefore,

$$k(t,s) = \sum_{i,j=1}^{\infty} < k, \Phi_{ij} > \Phi_{ij}(t,s). \tag{32}$$

Let

$$k_n(t,s) = \sum_{i,j=1}^{n} < k, \Phi_{ij} > \Phi_{ij}(t,s). \tag{33}$$

Then,

$$||k - k_n|| \to 0, \tag{34}$$

where $||\cdot||$ is the norm on $L_2([a,b] \times [a,b])$. Let K_n be the integral operator defined on $L_2([a,b])$ by,

$$(K_n f)(t) = \int_a^b k_n(t,s) f(s) \, ds. \tag{35}$$

So K_n is bounded, linear and of finite rank since $Im(K_n) \subset Sp\{\phi_1, \ldots, \phi_n\}$. Therefore by Lemma 1, K_n is also compact. Now,

$$||K|| \leq (\int_a^b \int_a^b |k(t,s)|^2 \, ds \, dt)^{1/2} = ||k||. \tag{36}$$

Thus applying (34) and (36) to $K - K_n$, we get,

$$||K - K_n|| \leq ||k - k_n|| \to 0. \tag{37}$$

So by Theorem A.1, K is compact.

To summarize, we have thus far shown that the integral operator K is bounded, linear, self adjoint and compact.

Although every linear operator on a finite dimensional Hilbert space over $\mathbb{C}$ has an eigenvalue, it is not true that even a self adjoint operator on an infinite dimensional Hilbert space must have an eigenvalue. The next few theorems are concerned with the eigenvalues of such operators.

Theorem A.2 *(a)Any eigenvalue of a self adjoint operator is real*

(b)If λ is an eigenvalue of $A \in L(H)$ then, $|\lambda| \leq ||A||$.

(c)If $A \in L(H)$ is compact and self adjoint then A has an eigenvalue and at least one of the numbers $||A||$ or $-||A||$ is an eigenvalue of A. ■

Theorem A.3 *Let $A \in L(H)$ be a compact self adjoint operator, where H is an infinite dimensional Hilbert space. Then the spectrum $\sigma(A)$ of A consists of zero and the eigenvalues of A. $(0 \in \sigma(A))$* ■

Now, $(\lambda I - K)$ is invertible for $\lambda \notin \sigma(K)$. Since K is compact and self adjoint, by Theorem A.3, $0 \in \sigma(K)$. Hence the inverse of $(\lambda I - K)$ is unbounded for $\lambda = 0$, which is equivalent to saying that the inverse of the integral operator K is unbounded.

From the above result, it is clear that in the infinite dimensional case the inverse problem is ill-posed in the sense of Hadamard since the inverse of the operator K is unbounded, solutions will not depend continuously upon the initial data.

Finite Matrix Approximation of the Convolution Kernel In the context of the tactile sensing problem, the convolution operator K has been approximated by a finite rank operator by discretizing the kernel k and considering only samples of the stress and strain. In order to justify the use of such an approximation, we state the following theorem.

Theorem A.4 *Every compact operator in $L(H_1, H_2)$ is the limit, in norm , of a sequence of operators of finite rank.* ■

In fact, in Section 2.4 it was shown that K is the limit (in norm) of the sequence of finite rank operators $\{K_n\}$.

If $\{\phi_j\}$ is an orthonormal basis for $L_2([a,b])$ we know that

$$\Phi_{ij}(t,s) = \phi_i(t)\bar{\phi_j}(s) \qquad i,j = 1,2,\ldots \tag{38}$$

forms an orthonormal basis for $L_2([a,b] \times [a,b])$. The infinite matrix representation, $[a_{ij}]$, of the integral operator K with respect to the basis $\{\phi_j\}$ is given by,

$$a_{ij} = < K\phi_j, \phi_i > = \int_a^b \int_a^b k(t,s)\phi_j(s)\bar{\phi_i}(t) \, ds \, dt = < k, \Phi_{ji} > . \tag{39}$$

Since $0 \in \sigma(K)$ and $A = [a_{ij}]$ is unitarily equivalent to K, zero is also in the spectrum of A and thus the inverse of A is also not a bounded operator. Theorem A.4 says that the operator K can be approximated arbitrarily well by an operator of finite rank, say $\widetilde{K}_n$ (of rank n) with finite matrix representation $\widetilde{A}_n$. As $n \to \infty$ the approximation becomes better, however, since $0 \in \sigma(K)$, as $n \to \infty$ at least one of the eigenvalues of $\widetilde{A}_n$ approaches zero. By Theorem 2(b), at least one of the eigenvalues of $\widetilde{A}_n$ approaches either $||K||$ or $-||K||$.

The condition number $P(A)$ of a matrix A is defined as

$$P(A) = \frac{\max_j |\lambda_j(A)|}{\min_j |\lambda_j(A)|} \tag{40}$$

In solving a finite system of linear equations of the form $Ax = y$, the condition number is a measure of sensitivity of solutions to errors in the initial data y and approximations made in the inversion of the matrix A (e.g. finite word length effects) on the solution x. Large values of $P(A)$ result in large errors in the solution. In the event that $P(A)$ is large ($P(A) = 1$ being the best case), we say that the matrix A is 'ill-conditioned'.

It is clear from the observations made earlier about the eigenvalues of $\widetilde{A}_n$ that as $n \to \infty$, $\widetilde{A}_n$ becomes increasingly ill-conditioned since the denominator in equation (40) approaches zero as the numerator approaches $||K||$.

We have shown that in the infinite dimensional case, the inverse problem is ill-posed in the sense of Hadamard. Ill-posedness of the problem in this case is due to the unboundedness of the inverse of the integral operator K which is in violation of Hadamard's third requirement that the solution depend continuously on the initial data for the problem to be well posed. We have justified approximating the operator K by a finite matrix since any compact operator in $L(H_1, H_2)$ can be approached (in norm) by a sequence of operators of finite rank. Lastly, we have made the observation that unboundedness of the inverse of K induces ill-conditioning of the finite matrix approximation to K.

B Energy Functions for Nonlinear RC Electrical Networks

Definitions Pertaining To Nonlinear Networks

Elements: Two terminal elements can in general be described by a relationship of the form,

$$f(i, \frac{di}{dt}, \frac{d^2i}{dt^2}, \ldots, v, \frac{dv}{dt}, \frac{d^2v}{dt^2}, \ldots, t) = 0 \tag{41}$$

Where v is the voltage across the two terminals of the element and i is the current through the element.

Nonreactive elements are those for which the dependence on time derivatives in equation (41) is absent, i.e. they can be described by $f(i, v, t) = 0$.

Time invariant elements are those for which the defining function is not explicitly time dependent

For time invariant nonreactive elements the locus of $f(i, v) = 0$ is called the *characteristic curve* of the element. For general time invariant elements, the i-v curve represents a trajectory in the phase space of the element viewed as a dynamical system.

A time invariant nonreactive element is said to be *passive* if the characteristic curve intersects the i-v axes at no point other than the origin. Otherwise the element is said to be *active*.

Sources: A current/voltage source is a time invariant active nonreactive element for which voltage/current is absent from the defining function $f(i, v)$.

Among the first theorems for linear electrical networks is Maxwell's Minimum Heat Theorem (1837) which is stated below in modern terminology.

Theorem B.1 (Maxwell's Minimum Heat Theorem) *For any linear, time-invariant, resistive network driven by voltage and/or current sources, of all the current distributions consistent with Kirchoff's current law, the only distribution which is also consistent with Kirchoff's voltage law and therefore the true distribution is the one which minimizes the quantity* $(W-2P_v)$ *where* W *is the total power dissipated by the resistors and* P_v *is the total power supplied by the voltage sources.* ■

Having chosen a set of generalized current coordinates say for example $i_1, \ldots, i_m$, Maxwell's theorem leads to the following set of m equations,

$$\frac{\partial}{\partial i_j}(W - 2P_v) = 0 \qquad j = 1, \ldots, m \tag{42}$$

which when solved for the current coordinates $i_1, \ldots, i_m$ determine all variables in the network.

The dual of Maxwell's theorem states that for the above network, among all voltage distributions which are consistent with Kirchoff's voltage law, the only one which is also consistent with Kirchoff's current law and therefore the true distribution is the one which minimizes $(W - 2P_I)$, where P_I is the total power supplied by the current sources. This dual theorem can be used to solve for all variable in the network by first solving for the generalized voltage coordinates.

Maxwell's theorem tells us that the quantities $(W - 2P_v)$ and $(W - 2P_I)$ are stationary with respect to the distributions of currents and voltages in the network respectively. It can be shown that these quantities are not in general stationary in the case of networks containing nonlinear elements. Hence to solve for voltages and currents in a nonlinear network in an analogous manner we must determine first the stationary quantities.

B.1 Invariants of Motion in Nonlinear Electrical Networks

Two quantities defined by Millar [37] relating to elements of a nonlinear network are the 'content' and the 'co-content' of an element.

Content and Co-content Let (i_1, v_1) be a point on the $i-v$ curve of a two terminal element. The 'content' of the element, denoted by G is defined by,

$$G = \int_0^{i_1} v \, di. \tag{43}$$

The 'co-content' of the element is denoted by J and is defined by,

$$J = \int_0^{v_1} i \, dv \tag{44}$$

We observe that for a passive element the total power dissipation is $W = J + G$ and that for a linear, passive element $J = G = W/2$

The total content (co-content) of a network is defined as the sum of contents (co-contents) of all constituent elements including current and voltage sources. We will use G and J to denote the total content and co-content respectively. We will use G_{jk} and J_{jk} to denote the content and co-content of the element connecting nodes j and k.

Stationary Quantities The following theorem identifies G and J as stationary quantities.

Theorem B.2 (Millar) *If in an active (possibly reactive) network, the sum J of the co-contents of all the constituent elements is expressed in terms of the defining number of voltage coordinates of the network subject only to the restrictions of Kirchoff's voltage law, then J is stationary for the actual distribution of voltages.* ■

The dual of this theorem is obtained by replacing co-content by content, J by G, and voltage by current everywhere.

Theorem B.2 provides us with the equivalent of Maxwell's theorem for nonlinear networks of time invariant elements. To restate the above theorem in a manner analogous to Maxwell's theorem, we can say;

In any active (possibly reactive) network, of all distributions of current/voltage consistent with Kirchoff's current/voltage law, the ones for which G/J is stationary are the only ones that are also consistent with Kirchoff's voltage/current law and are thereby the true distributions.

Thus if we express G/J in terms of the defining number (m/n) of current/voltage coordinates, we can determine all currents and voltages in the network by solving either one of the following sets of simultaneous partial differential equations, $\frac{\partial G}{\partial i_r} = 0 \qquad r = 1, \ldots, m$, or $\frac{\partial J}{\partial v_q} = 0 \qquad q = 1, \ldots, n$.

Invariants of Motion The next theorem identifies G and J as invariants of motion. That is, as a network evolves in time following an impulsive change in one or more of the current or voltage sources, the total content and the total co-content of the network are conserved.

Theorem B.3 (Millar) *In any network of time invariant elements (possibly including sources), the total content G and the total co-content J are invariants of motion. (i.e. $\frac{dG}{dt} = 0$ and $\frac{dJ}{dt} = 0$, where $\frac{dG}{dt}$ and $\frac{dJ}{dt}$ are the total time derivatives of G and J given by,*

$$\frac{dG}{dt} = \frac{\partial G}{\partial t} + \frac{\partial G}{\partial i_1}\frac{di_1}{dt} + \cdots + \frac{\partial G}{\partial i_m}\frac{di_m}{dt}$$

and

$$\frac{dJ}{dt} = \frac{\partial J}{\partial t} + \frac{\partial J}{\partial v_1}\frac{dv_1}{dt} + \cdots + \frac{\partial J}{\partial v_n}\frac{dv_n}{dt}$$

where m is the defining number of generalized current coordinates and n is the defining number of generalized voltage coordinates. ■

In a directed network with b branches and m nodes, the set of branch currents $i = (i_1, ..., i_b)$ and the set of branch voltages $v = (v_1, ..., v_b)$ are vectors in a b-dimensional Euclidean vector space $\mathcal{E}^b$ with the inner product defined by $< x, y >= \sum_{\mu=1}^{b} x_\mu y_\mu$. Let $\mathcal{I}$ be the set of all vectors in $\mathcal{E}^b$ such that if $i \in \mathcal{I}$ then the constraint of Kirchoff's current law is satisfied for each node in the network, i.e. $\sum_{node} i_\mu = 0$. Similarly, let $\mathcal{V}$ be the set of all vectors in $\mathcal{E}^b$ such that if $v \in \mathcal{V}$ then Kirchoff's voltage law is satisfied, i.e $\sum_{loop} v_\mu = 0$. $\mathcal{I}$ and $\mathcal{V}$ are clearly subspaces of $\mathcal{E}^b$ since they are defined via linear relationships (Kirchoff's Laws). The following theorem which appears in [2] is easily obtained using Tellegen's Theorem (see [49]) which states that $\mathcal{I}$ and $\mathcal{V}$ are orthogonal subspaces of $\mathcal{E}^b$.

Theorem B.4 (Brayton–Moser) *Let Γ denote a one dimensional curve in $\mathcal{I} \times \mathcal{V}$ with coordinates denoted by i and v. Then,*

$$\int_\Gamma \sum_{\mu=1}^{b} v_\mu di_\mu = \int_\Gamma \sum_{\mu=1}^{b} i_\mu dv_\mu = 0 \qquad \blacksquare \tag{45}$$

Having stated the above two theorems and having defined the complete set of variables $v* = (v_1, \ldots, v_N)$ for the network, we proceed with the proof of Theorem 3.1

Proof Of Theorem 3.1: From Theorem B.4 we know that $\int_\Gamma \sum_{\mu=1}^{b} i_\mu dv_\mu = 0$. We choose Γ from a fixed initial point to a variable end point in $\mathcal{E}^b$ such that along Γ the characteristic relationships of the constituent elements of the network are satisfied. We can write equation (45) in the following form:

$$\int_\Gamma \sum_{\rho=1}^{N} i_\rho dv_\rho + \int_\Gamma \sum_{\mu=N+1}^{b} i_\mu dv_\mu = 0 \tag{46}$$

The first integral is over all capacitive branches and the second is over all other branches. Note that if the first integral on the left is independent of the path Γ then the integration and summation may be interchanged to obtain $-P(v)$ as defined in the statement of Theorem 3.1. Let,

$$\widetilde{P}(v) = -\int_\Gamma \sum_{\rho=1}^{N} i_\rho dv_\rho. \tag{47}$$

Let $\xi = \sum_{\rho=1}^{N} i_\rho dv_\rho$. For the integral in (47) to be independent of Γ, it is necessary that ξ be a perfect differential. That is, we can write ξ as, $\xi = d\sigma$, where $\sigma = \sigma(v_1, \ldots, v_N)$ and,

$$d\sigma = \frac{\partial \sigma}{\partial v_1} dv_1 + \cdots + \frac{\partial \sigma}{\partial v_N} dv_N. \tag{48}$$

But since we want $\xi = d\sigma$ we must have,$d\sigma = i_1 dv_1 + \cdots + i_N dv_N$. Equivalently, we need $i_\rho = \frac{\partial \sigma}{\partial v_\rho} \qquad \rho = 1, \ldots, N$, which is the case *if and only if*,

$$\frac{\partial i_\rho}{\partial v_\eta} = \frac{\partial^2 \sigma}{\partial v_\rho \partial v_\eta} = \frac{\partial i_\eta}{\partial v_\rho} \quad \eta, \rho = 1, \ldots, N. \tag{49}$$

By hypothesis *(H2)* (49) holds. Hence $\widetilde{P}(v)$ is a function of the endpoints of Γ alone, and choosing the origin as a starting point for Γ, we have,

$$\widetilde{P}(v) = P(v) = -\sum_{n=1}^{N} \int_0^{v_n} i_n dv_n. \tag{50}$$

In this case we have,

$$i_\rho = -\frac{\partial P(v)}{\partial v_\rho} = C_\rho \frac{dv_\rho}{dt} \qquad \rho = 1, \ldots, N, \tag{51}$$

where the equality on the right is obtained by the dynamical law of capacitors.

Therefore,

$$C_\rho \frac{dv_\rho}{dt} = -\frac{\partial P(v*)}{\partial v_\rho} \quad \rho = 1, \ldots, N \tag{52}$$

where $v* = (v_1, \ldots, v_N)$. We now write the system of differential equations defining the dynamical behavior of the network in the following vector form,

$$-C\dot{v^*} = \frac{\partial P(v^*)}{\partial v^*} \tag{53}$$

where $C = diag(C_1, ..., C_N)$, and $\frac{\partial P(x)}{\partial x}$ is the gradient of $P(x)$. Since the matrix C is positive definite and symmetric, we know that $\frac{d}{dt}P(v^*(t)) \leq 0$ and $\frac{d}{dt}P(v^*(t)) = 0$ if and only if v^* is an equilibrium of the gradient system (53). Hence if $\widetilde{v^*}$ is an isolated minimum of $P(v^*)$ then $\widetilde{v^*}$ is an asymptotically stable equilibrium of the gradient system (53). Therefore the equilibrium states of the network correspond to stationary points of $P(v^*)$, which we shall call the energy function, and the local minima of $P(v^*)$ are the stable equilibria of the network. If in addition to this $P(x) \rightarrow \infty$ as $||x|| \rightarrow \infty$ then it can be shown using a well known result from Lyapunov stability theory that all solutions to (53) approach one of the set of equilibrium solutions as $t \rightarrow \infty$. ■

We observe that $P(v*)$ is just the negative of the co-content of *all independent capacitive branches* in the network. The total co-content of the network, J, is an invariant of motion (Millar [37]) even for dissipative systems where the total energy is not conserved. The analogy to kinetic and potential energy of a nondissipative mechanical system is evident. $P(v*)$ can be regarded as a type of potential energy which is minimized as the system settles. The sum of the co-contents of all other branches in the network plays the role of kinetic energy. J, which is the sum of these two quantities is like the total energy of the system and is conserved.